Name – Spursh Ujjawal

SJSUID - 010471071

Ques – 22

Link - <http://jmcauley.ucsd.edu/data/amazon/links.html>

From the Dashboard select the Data & Analytics.

Click on Analytics and create New Instance and New Notebook.

I’m using Scala for my Notebook.

Drag and drop the data file to Data Source.

Run the following queries:

* import org.apache.spark.sql.functions.\_
* val sqlContext = new org.apache.spark.sql.SQLContext(sc)
* val df = sqlContext.read.json("swift://notebooks.spark/reviewsMusic.json.gz2")
* df.sort(desc("overall")).show(10)

Selects the top 10 Music records from the data set given .